

Tartu Ülikool

Matemaatika- informaatikateaduskond

Arvutiteaduse instituut

Gerth Jaanimäe

Infotehnoloogia

Eesti framenet'i viimine MySQL'i

Bakalaureusetöö (6 EAP)

Juhendaja: Neeme Kahusk

Seminarijuhendaja: Mare Koit

Tartu 2014

Eesti framenet'i viimine MySQL'i

Käesoleva bakalaureusetöö eesmärk on viia Eesti freimileksikon üle MySQL'i. See võimaldab veebiliidese abil süsteemi kiiremat ja mugavamat kasutamist. Framenet on selline leksikaalsemantiline ressurss, kus sõnad on erinevalt teistest jaotatud freimidesse. Freimiks nimetatakse antud juhul olukorra, sündmuse või objekti kirjeldust.

Võtmesõnad: Framenet, freim, lekseemid, elemendid, suhted.

Transferring the Estonian framenet into MySQL

The aim of the thesis is to transfer Estonian framenet into MySQL, in order to use it more comfortably and efficiently via web interface. Framenet is a kind of lexical semantic resource, where the words are put in frames, which are in this case descriptions of the situation, event or an object.

Keywords: Framenet, frame, lexical units, elements, relations.

Table of Contents

Sissejuhatus	4
1. Framenet'i kirjeldus	5
1.1. Freim, elemendid ja lekseemid	5
1.2. Framenet'i ja teiste semantiliste sõnastike võrdlus	5
1.3. Framenet'i kasutusala	6
1.4. Freimide omavahelised suhted	6
1.4.1. <i>Inheritance</i> – pärilus	6
1.4.2 <i>Perspective on</i> (perspektiiv)	7
1.4.3 <i>Sub frames</i> - Alamfreimid	7
1.4.4. <i>precedes</i> - Eelneb.....	7
1.4.5. <i>Causative of</i> ja <i>Inchoative of</i> - kausatiiv ja inhoatiiv	7
1.4.6. <i>Using</i> - Kasutab.....	7
1.4.7. <i>See also</i> vaata ka.....	8
1.5. Semantilised tüübid.....	8
1.5.1. <i>Ontological types</i> - Ontoloogilised tüübid.....	8
1.5.2. <i>Framal type</i> – freimi tüübid	8
1.5.3. Leksikaalsed tüübid	9
2. Tarkvara realiseerimine	10
2.1. Tarkvara nõuded	10
2.2. Lahenduse eelised eelkäijate ees.....	10
2.3. Sarnased realisatsioonid.....	11
2.4. Tarkvara loomisel tehtud sammud:	11
2.5. Framenet'i andmebaasi kirjeldus.	12
2.6. Realiseerimiseks valitud tehnoloogiad	13
2.7. Võimalikud lisandused.....	14
3. Tarkvara paigaldamine ja kasutamine.....	15
3.1. Tarkvara paigaldamine	15
3.2. Tarkvara kasutamine.	15
Kokkuvõte.....	18

Viited	19
Summary	20

Sissejuhatus

Käesolev bakalaureusetöö käsitleb eesti framenet'i viimist MySQL'i. Framenet'iks nimetatakse sellist leksikaalsemantilist ressursi, kus erinevalt tavalistest, on sõnad jaotatud freimidesse ehk olukorra kirjeldustesse. Sellist sõnastikku saab kasutada mitmesugusel viisil loomuliku keele uurimiseks. Kuigi eesti freimileksikon on juba olemas xml formaadis, on selle mugavamaks haldamiseks tarvis luua andmebaas. Lisaks sellele proovitakse antud bakalaureusetöö raames luua andmebaasile kasutajaliides.

Antud bakalaureusetöö koosneb kolmest osast.

Esimeses peatükis tuleb juttu sellest, mis on framenet, freim, elemendid, freimide omavahelised suhted jne. Teine peatükk käsitleb tarkvara realiseerimist, mida tarkvara loomisel kasutati ning miks antud lahendus on oma eelkäijatest parem.

Lisaks on seal välja toodud ka andmebaasi lühike kirjeldus. Kolmandas peatükis on kirjas tarkvara lühike paigaldamis- ja kasutusjuhend. Sellele järgneb nii eesti kui ka inglisekeelne kokkuvõte. Lisaga on kaasa pandud andmebaasi eksport- ja php koodifailid.

1. Framenet'i kirjeldus

Käesolevas peatükis kirjeldatakse framenet'i struktuuri ja selle erinevaid osasid. Tutvustatakse lugejale, mis on lekseem, freim, selle elemendid ning kirjeldatakse põgusalt erinevaid freimide omavahelisi suhteid ja semantilisi tüüpe.

1.1. Freim, elemendid ja lekseemid

Framenet'iks nimetatakse semantilist sõnastikku, kus sõnad on paigutatud freimidesse, ehk teisisõnu olukorra, sündmuse või ka objekti kirjeldustesse.

Näiteks freim „kuumust rakendama“ (ingl. K. *apply heat*) kirjeldab olukorda, mis hõlmab endas kokka, sööki ja eset, millega kuumutatakse. Sealhulgas on freimi esilekutsujaks sõnad nagu küpsetama, praadima, keetma. Eelkirjeldatud rolle aga nimetatakse freimi elementideks ning sõnu, mis freime välja kutsuvad lekseemideks, mõnikord edaspidi ka sõnadeks. Osad freimid on abstraktsemad, nagu *Change_position_on_a_scale* (skaalalasukohta muutma), mille kutsuvad esile sõnad nagu vähenema, vähendama, suurenema, alla kukkuma, tõusma jne ning elementideks *Item* (ese), *Attribute* (atribuut), *Initial.value* (algne väärtus) ning *Final.value* (lõppväärtus). Lihtsamal juhul on freimi esilekutsujaks tegusõna ning selle elemendid on sellest semantiliselt sõltuvad. Neid saab jaotada ka kohustuslikeks (*core elements*) ja mittekohustuslikeks. Kohustuslikud elemendid peavad freimi esilekutsuva sõnaga alati kaasas käima. [1]

Üks mitmetähenduslik sõna kuulub mitmesse erinevasse semantilisse freimi. Näiteks lemma küpsetama kuulub kahte freimi *apply_heat* ja *Cooking_creation*. Näitelause: Michelle küpsetas kartuleid 45 minutit. Michelle küpsetas emale sünnipäevaks koogi. Need tekitavad kaks erinevat lekseemi erinevate definitsioonidega. [1]

1.2. Framenet'i ja teiste semantiliste sõnastike võrdlus

Iga sõna on ühendatud semantilise freimiga ja seega ka teiste sõnadega, mis seda freimi esile kutsuvad. See teeb framenet'i sarnaseks tesaurusega, mis rühmitab kokku sarnase tähendusega sõnad.

Wordnet ja ontoloogiad võimaldavad mingit tüüpi hierarhiat oma n.ö. sõlmede vahel (ingl. K. *Node*). Nii on ka framenet'i puhul, mis sisaldab suhete võrgustikku erinevate freimide vahel. Suhteid on mitut erinevat tüüpi, nendest lähemalt peatükis 1.4.

Kuna freimidesse ei kuulu tavaliselt nimisõnad, ei ole framenet eriti otstarbekas asjade ontoloogia.

Framenet'i analüüs käib tavaliselt freimide kaupa, samas kui wordnet'i analüüs käib lemmade kaupa tähestikulises järjekorras. Seega, kui tavalistes sõnastikes mõõdetakse arengut sõnade arvuga, siis framenet'is on oluline freimide hulk. Kui üks sõna on paigutatud mitmesse freimi, ei tähenda see seda, et see tulevikus veel mõnda freimi kuuluda ei võiks. [1]

1.3. Framenet'i kasutusalad

Loomuliku keele töötluste uurijad saavad framenet'i kasutada info eraldamiseks, masintõlkeks, juhtumi tuvastamiseks, sentimendi analüüsiks. Keeleteadlastele pakub see detailset infot sõnavara erinevatest kombineerimisvõimalustest. [2]

Inglise framenet'i andmeid on kasutatud näiteks järgmisteks projektideks: Poliitiliste kampaaniate säutsude analüüsimiseks, loomuliku keele mitmesuguste alade uurimiseks jne. [3]

1.4. Freimide omavahelised suhted

Freimide omavahelisi seoseid iseloomustavad mitmed erinevad suhted. Allpool on välja toodud nende inglisekeelne nimetus koos eestikeelse tõlkega ja lühike kirjeldus.

1.4.1. *Inheritance* – pärilus

See on tugevaim suhe freimide vahel, kus kõik emafreimi (*parent frame*) omadused on tõesed ka tütarfreimi (*child frame*) puhul. Viimasele kehtivad emafreimi elemendid, enamus semantilistest tüüpidest ja tema suhted teiste freimidega. Mitte rangelt semantilised omadused nagu sõnad, mis freimi esile kutsuvad või *see also* (vaata ka) suhe teiste freimidega ei ole pärilikud. [1]

1.4.2 Perspective on (perspektiiv)

See suhe näitab vähemalt kahe erineva vaatepunkti olemasolu neutraalse freimi suhtes. Näiteks freimis *Measured_scenario*, kus kirjeldatakse kellegi või millegi atribuudi väärtust, saab olukorda vaadata nii täpse mõõtmise vaatenurgast, „Joey kaalus seitse naela“, kui ka suhtelise „Joey oli raske“. Elemendid on mõlemal juhul erinevad ja seega ei tohiks sõnad olla samas freimis, aga nad viitavad sarnasele olukorrale. Antud suhe võimaldab seega kahte freimi omavahel ühendada. Veel üheks näiteks on freimid *Get_a_job* (tööd saama) ja *Hiring* (palkama) on seotud freimiga *Begin_employment* (tööle võtmine). [1]

1.4.3 Sub frames - Alamfreimid

Leidub keerukaid freime, mis viitavad seisundite või üleminekute järjenditele. Nendest igaüht annaks kirjeldada eraldi freimina. Keerukama freimi elemente võib ühendada ka alamfreimide omadega, aga mitte alati. Selles mõttes vastandub see pärilussuhtega (vt 1.2.1.). On oluline tähele panna, et alamfreim võib olla ise samuti keerukas ning seega olla sellel veel alamfreime. [1]

1.4.4. precedes - Eelneb

See suhe kehtib ainult kahe keeruka freimi komponentfreimi vahel. See on ainus suhe, kus tsüklid on lubatud. Näiteks une ja ärkveloleku freimis eelneb ärkvel olemine magama jäämisele, see omakorda eelneb magamisele, see omakorda ärkamisele ning see jällegi esimesele freimile, ehk ärkvel olemisele. [1]

1.4.5. Causative of ja Inchoative of - kausatiiv ja inhoatiiv

Antud suhe iseloomustab semantiliselt eriti lähedasi, aga mitte pärilike omadustega freime. Näiteks freimid: *Position_on_a_scale*, *Change_position_on_a_scale* ja *Cause_change_of_scalar_position*. [1]

1.4.6. Using - Kasutab

Üks kindel freim viitab sageli väga üldiselt mõnele teisele abstraktsele skemaatilisele freimile. Vastavat suhet kasutatakse peamiselt ainult siis, kui vaid

osa tütarfreimist viitab mõnele teisele. Näiteks freim *Volubility* (paljusõnalisus) kasutab freimi *communication* (suhtlemine), kuna esimene kirjeldab suhtlemisel toimuvaid sündmusi. Üks freim võib kasutada mitut freimi. Näiteks kasutab freim *Judgment_communication* nii freimi *Judgment* kui ka *Statement*. Esimese freimi omadusi ei päri see selle pärast, et pole kognitiivse seisundi alamtüüp. Teise freimi omadusi ei pärita, sest selle element *Message* (sõnum) jaotatakse kahe elemendi vahel *Evaluee* (hindaja) ja *Reason* (põhjus). [1]

1.4.7. See also vaata ka

Juhul kui leidub rühm sarnaseid freime, mida on vaja hoolikalt omavahel eraldada, võrrelda ja eristada, on igal sellisel freimil vaata ka suhe vastava freimiga. Antud suhe ei hõlma endas freimi elemente ega seotud freimide alamfreime. [1]

1.5. Semantilised tüübid

Semantilisi tüüpe kasutatakse info, mida freimi elemendid ja nende omavahelised suhted ei kirjelda, talletamiseks. Näiteks semantilised tüübid "*Positive_judgement* (positiivne hinnang)" ja "*Negative_judgement* – negatiivne hinnang", mis näitavad kõneleja suhtumist olukorda, saab rakendada lekseemidele mitmes erinevas freimis.

Allpool on välja toodud erinevad semantilised tüübid ja nende kirjeldused.

1.5.1. Ontological types - Ontoloogilised tüübid

Neid semantilisi tüüpe kasutatakse sõnade, freimide ja nende elementide otseste tähenduste liigitamiseks. Juhul, kui freimile on määratud selline tüüp, rakendub see automaatselt sinna kuuluvatele lekseemidele. Freimi elementide korral näitab see aga seda, millist nii-öelda täidet selle puhul oodatakse, mitte seda, mis rolli see omandab. [1]

1.5.2. Framal type – freimi tüübid

Neid saab rakendada ainult freimidele ja mitte lekseemidele ega seotud freimidele. Sinna hulka kuuluvad:

- *Non lexical frame* – mitte leksikaalne freim – sinna kuuluvad freimid ei sisalda ühtegi lekseemi ja on vaid selleks, et ühendada mitut erinevat freimi.

Non-perspectivalized_frame – Kasutatakse freimide puhul, kuhu kuulub palju erinevaid, kuid sarnase taustaga lekseeme. Sellistel freimidel puuduvad kindlad elemendid, sündmustele määratud aeg ja osalised. Neid saaks eraldada mitmeteks väiksemateks freimideks, aga need sisaldaksid liiga vähe lekseeme. [1]

1.5.3. Leksikaalsed tüübid

- *Transparent_nouns* (läbipaistvad nimisõnad)

Erinevalt paljudest teistest nimisõnadest, on nende funktsiooniks kirjeldada mingit muud nimisõna. Näiteks:

Ma lugesin seda tüüpi raamatut koolis. Pane see riuli vasakule poolele.

- *Bound_LU* (seotud lekseemid)

Osasid sõnu saab kasutada ainult siis, kui need on koos mõnede õigete teiste sõnadega.

- *Biframal_LU* (kahefreimiline lekseem)

Lekseemidel on tavaliselt tähendus, mis on freimi tähenduse alamtüüp.

Kahefreimiline semantiline tüüp aga näitab, et sõna ja freimi semantika on küll sarnased, aga mitte täpselt samad. Sellistel juhtudel oleks võimalik tekitada ka eraldi väikesed freimid, aga aja ja ressursi kokkuhoiu mõttes ei oleks see eriti mõttekas.

- *Affect_describing: Positive- and Negative.judgement (positiivne ja negatiivne arvamus)*

Paljud lekseemid kirjeldavad ilmselt või kaudselt positiivset ja negatiivset arvamust, olemata sama tüüpi. Näiteks sõna meeldima, mis asub freimis *Experiencer_subj.* ja torkav freimis *Stingyness*. See semantiline tüüp võimaldab selliseid sõnu omavahel võrrelda, vaatamata sellele, et nad eriti sarnased ei ole. [1]

2. Tarkvara realiseerimine

Eesti framenet'i mugavamaks haldamiseks viidi andmed üle MySQL'i. Selle veebis kasutamiseks loodi php-s kasutajaliides. Järgnevates jaotistes kirjeldatakse tarkvara nõudeid, realiseerimist, sarnaseid lahendusi ning põhjendatakse kasutatud tehnoloogiate valikuid.

2.1. Tarkvara nõuded

Funktsionaalsed nõuded:

- * Süsteem peab kasutajale kuvama freimi ja sõnade kohta käivat teavet.
- * Kasutaja peab saama freime ja sõnu otsida või esitähed järgi sirvida.
- * Volitatud kasutajad peavad saama lisada, muuta ja kustutada sõnu, freime, nende omavahelisi suhteid ja elemente.

Mittefunktsionaalsed nõuded:

- Tarkvara peab olema kättesaadav 95% ajast.
- Tarkvara peab olema kasutatav enamlevinud brauseritega
- Tarkvara peaks kasutatav olema ekraanilugemistarkvaraga
- Süsteemi turvaklass ISKE: s0k2t1 andmed on avalikud, süsteem peab olema piisavalt käideldav ning andmete rikkumine või hävitamine põhjustab piisavalt ebameeldivusi)

2.2. Lahenduse eelised eelkäijate ees

Enne antud tarkvara loomist oli eesti freimileksikon küll olemas, aga see oli xml formaadis. [4]

Seega eeldas selle kasutamine, s.t. erinevate toimingute (sõnade ja freimide info vaatamine, muutmine, kustutamine) kasutajatelt infotehnoloogiaalaseid teadmisi

ja oli üsnagi aeganõudev. MySQL'i kaudu php, ehk siis veebi vahendusel on seda kõike oluliselt kiirem ja mugavam teha.

Lisaks puudusid xml formaadis leksikonis freimide definitsioonid ja omavahelised suhted ning sõnade tähendused ja näitelaused.

2.3. Sarnased realisatsioonid

Eesti realisatsioonidest on teatud määral framenet'ile sarnane wordnet. [5] Sellest saab veidi lugeda peatükis 1.2). Viimases on sõnad esitatud sünohulkadena, ehk siis teatud sõna juurde on märgitud tema sünonüümid. Eesti wordnet'i andmeid on kasutatud ka sõnade tähenduste ja näitelausete saamiseks.

Inglisekeelsetest realisatsioonidest on antud lahenduse eeskujuks inglise keelele loodud framenet. [6] Sealt on võetud suures osas eeskuju andmebaasi tabelite loomiseks. Samuti on sealt imporditud freimide definitsioonid ja freimide omavahelised suhted.

2.4. Tarkvara loomisel tehtud sammud:

- Andmebaasi loomine
- Tabelite loomine vastavate väljadega
- Xml formaadis eesti freimileksikoni andmebaasi importimine. Kasutati programmeerimiskeeles python kirjutatud skripte. Andmebaasi loeti sisse sõnad, freimid ja nende elemendid.
- Freimide definitsioonid ja omavahelised suhted. Imporditi inglise framenet'i andmetest, mis olid samuti xml-keeles, samal meetodil, mis eelmises punktis.
- Sõnade tähendused võeti getcom formaadis olevatest wordneti andmetest samuti Pythonis kirjutatudskripti abil.
- Loodi andmebaasi vastavad kasutajad
- Seejärel loodi andmebaasile php-s kasutajaliides
- Tarkvara testimine (brauseritega firefox ja google crome)

2.5. Framenet'i andmebaasi kirjeldus.

Andmebaasi struktuur on üldiselt üsna lihtne ja robustne. Sellesse kuuluvad järgmised tabelid. Kuna sql käsud on inglise keeles, on üldise loetavuse huvides tabelite ja väljade nimed samuti inglisekeelsed:

Tabel frame:

Id, name, definition

Tabel word:

Lexical_unit, frame, meaning, example

Frame on seotud tabel frame veeruga id.

Tabel elements:

Id, frame, name, optional

Frame on seotud tabel frame veeruga id.

Tabel relation:

Frame1, relation_type, frame2

Frame1 ja frame2 on seotud tabel frame veeruga id.

Kasutajad ja nende õigused:

Tavakasutaja framenet (select)

Volitatud kasutaja admin (select, update, insert, delete).

2.6. Realiseerimiseks valitud tehnoloogiad

Antud tarkvara loomiseks kasutati php-d ja MySQL-i. Valik sai tehtud järgmistel kaalutlustel:

MySQL'i eelised:

- avatud lähtekood, tänu millele saab seda tasuta alla laadida ja kasutada.
- Robustsus – Hea jõudlus tänu MyISAM'i kasutamisele.
- Võtab väga vähe kettaruumi.
- erinevate suuremate operatsioonisüsteemide tugi
- Kergesti õpitav
- Piisaval hulgal võimalusi. [7]

PHP eelised:

- Avatud lähtekood
- Töötab erinevatel operatsioonisüsteemidel
- Kasutajasõbralikkus
- Kiirus -
- Laiendused – Kuna php on avatud lähtekoodiga, on erinevad teegid ja laiendused kergesti kättesaadavad.
- Kerge paigaldatavus
- Suur kasutajatekommuun
- Turvalisus – Erinevaid turvalisuse parameetreid saab hõlpsalt muuta .ini failis.
- Ühildub hästi apache ja MySQL'iga. [8]

2.7. Võimalikud lisandused

Kuigi süsteemi põhifunktsionaalsus töötab, pole ükski lahendus täiuslik. Nii saaks ka sellele lahendusele lisada veel mitmeidki funktsioone, mis antud bakalaureusetöö skoobist välja jäid. Nendeks võiksid olla:

- Võimalus andmebaasi eksportida xml formaati. Teksti märgendamiseks oleks see tõenäoliselt lihtsam. Teine võimalus oleks kirjutada skript, mis andmebaasist andmeid pärib.
- Sõnade muutmisel või lisamisel võiks tähendused ja näitelaused kätte saada otse wordnetist. Lisaks oleks võimalik võrreldes omavahel freimis ja wordneti sünohulgas olevaid sõnu määrata vaikimisi variant, mis oleks tõenäoliselt õige. Selle probleem seisneb selles, et wordnet on getcom formaadis. ning kuna seda uuendatakse, siis on sealt kõige värskemad infot suhteliselt keeruline kätte saada. Üks võimalus oleks vastavad andmed iganädalaselt allalaadida ning siis viia need sobivale kujule.

3. Tarkvara paigaldamine ja kasutamine

Käesolev peatükk annab lühikese õpetuse, kuidas tarkvara paigaldada ja kasutada.

3.1. Tarkvara paigaldamine

Tarkvara käivitamiseks peavad eelnevalt olema paigaldatud:

- Linuxis Apache2, php5, libapache2-mod-php5, MySQL-server, libapache2-mod-auth-MySQL, php5-MySQL.
- Windowsis sobib selleks nt xampp

Järgnevad sammud eeldavad kasutajal MySQL'i ja apache algtasemel tundmist.

Andmebaasi paigaldamine:

1. Luua andmebaas framenet.
2. . Kasutajad järgmiste õigustega: (parooli saab vaadata ja vajadusel muuta failis users.php)

Framenet (tavakasutaja) – select

Admin – select, insert, update, delete.

3. Importida andmed andmebaasi failist framenet.sql

Php failid tuleb kopeerida oma veebiserveri vastavasse kataloogi. Apache'l on selleks näiteks /var/www.

3.2. Tarkvara kasutamine.

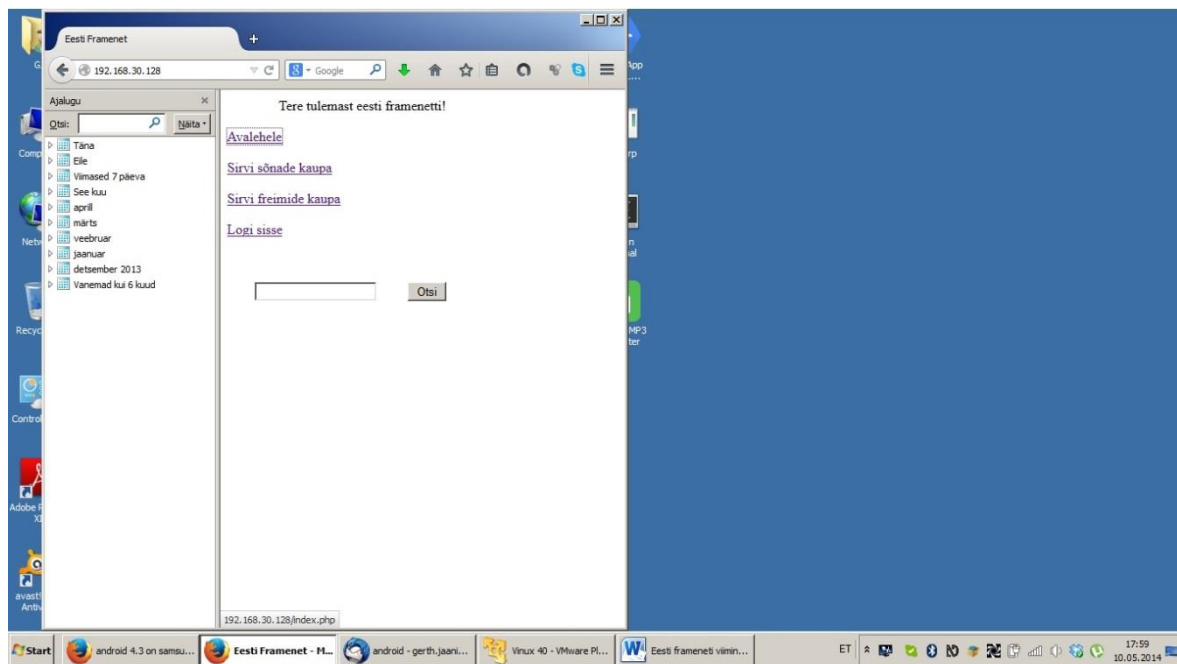
Tarkvara kasutamiseks tuleb sisestada oma brauseri aadressiribale kas vastava lehekülje või masina aadress või kui see on paigaldatud enda arvutisse, siis localhost.

Avalehel avanevad järgmised valikud:

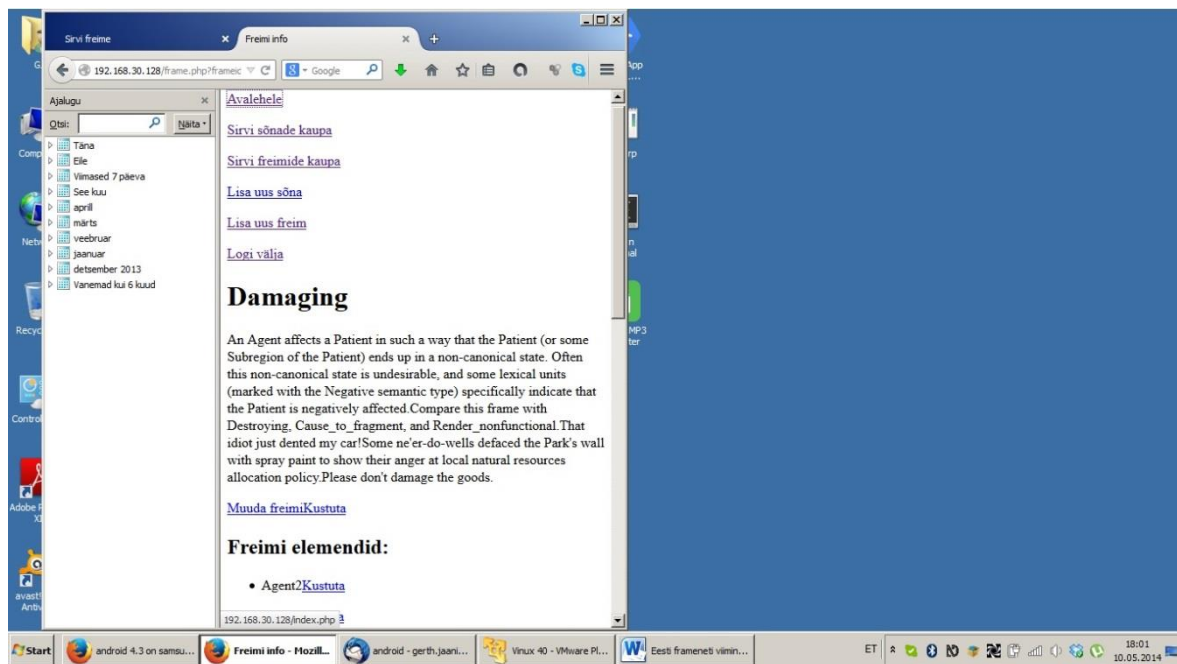
Sirvi sõnade kaupa, sirvi freimide kaupa ja logi sisse. Lisaks saab seal teostada ka freimide ja sõnade otsingut.

Selleks, et freimide ja sõnade kohta käivat teavet muuta, peab olema sisse logitud.

Pärast seda ilmuvad sõnade ja freimide sirvimisel vastava üksuse juurde lingid muuda ja kustuta. Lisaks saab neid toiminguid sooritada ka freimi infot vaadates.



Avaleht



Freimi info sisselogitud kasutajale

Kokkuvõte.

Käesoleva bakalaureuse töö raames viidi eesti freimileksikon xml formaadist üle MySQL'i ning loodi sellele php-s kasutajaliides. See on oluline eelkõige sellepärast, et võimaldab süsteemi igal pool senisest mugavamalt kasutada ja hallata.

Töö käigus tekkisid kõige suuremad raskused andmete andmebaasi üleviimisel. Kuigi antud bakalaureusetöö kirjalikus osas käsitletakse seda üsna vähe, oli see ajaliselt kõige mahukam. Seda peamiselt sellepärast, et erinevad üleviidavad osad olid erineval kujul (vt 2.4) ning nende tundma õppimine nõudis päris palju aega. Lisaks, kuna polnud eriti palju kogemusi pythoni mitmesuguste stringitöötlus meetoditega, võttis nende erinevate nüansside tundmaõppimine samuti ajalist ressursi. Seevastu osutus üsna huvitavaks php-s kasutajaliidese kirjutamine.

Eesti framenet'i andmebaasi loomine oli üks väike samm edasi eesti keele uurimises, kuna seda kasutatakse loomuliku keele uurimiseks mitmesugusel erineval viisil. Eelkõige oleks see kasulik peamiselt lingvistidele, aga seda saaks kasutada ka keele huvialase uurimise otstarbel.

Viited

1. Josef Ruppenhofer Michael Ellsworth Miriam R. L. Pet ruck Christopher R. Johnson Jan Scheffczyk - FrameNet II: Extended Theory and Practice

2. About FrameNet – viimati vaadatud 21.04.2014

<https://framenet.icsi.berkeley.edu/fndrupal/about>

3. Framenet users – viimati vaadatud 10.05.2014

https://framenet.icsi.berkeley.edu/fndrupal/framenet_users

4. Eesti freimileksikoni xml fail – viimati vaadatud 10.05.2014

<http://www.keeletehnoloogia.ee/ekt-projektid/semantika-vahendid-eesti-keelele/freimileksikon>

5. Eesti wordnet – Viimati vaadatud 12.05.2014

<http://www.cl.ut.ee/ressursid/teksaurus/>

6. Inglise keelele loodud framenet – Viimati vaadatud 11.05.2014

<https://framenet.icsi.berkeley.edu/fndrupal/home>

7. What are the advantages of MySQL? - Viimati vaadatud 22.04.2014

<http://www.rapidprogramming.com/questions-answers/What-are-the-advantages-of-MySQL--574>

8. The advantages of php - Viimati vaadatud 2.05.2014

<http://www.tutorialchip.com/php/advantages-of-php-programming/>

Summary

Transferring Estonian framenet into MySQL

Bachelore thesis

Gerth Jaanimäe

Abstract

The aim of the thesis was to transfer Estonian framenet into MySQL and write a php web interface for it. Framenet is a kind of dictionary, or to be more precise, lexical resource, which unlike many, groups words (often also described as lexical units) together by frames, which in this case are descriptions of the situation, event or an object. Frames are connected to each other by different relation types. Besides lexical units, each frame has also different elements.

The Estonian frame lexicon was previously in xml format, so it's usage was quite time consuming and it required the knowledge of that language. Also, there were no frame definitions or relations of the frames. The meanings and examples of lexical units were taken from Estonian wordnet and frame definitions and relations were imported from the framenet for English language.

The software that was written allows users to search, browse and view information about the lexical units and frames more comfortably. Authenticated users can also add, modify and delete them.

The thesis consists of three parts. The first chapter describes the structure of framenet and different parts of it. The second chapter describes the implementation of the software, including the description of the database. The third and final chapter is the instruction for installing and using the software.

The main usages of the framenet are different kinds of research of natural language, sentiment analysis and machine translation.

Lihtlitsents lõputöö reprodutseerimiseks ja lõputöö üldsusele kättesaadavaks tegemiseks

Mina _____

Gerth Jaanimäe

sünnikuupäev: 21.06.1989

annan Tartu Ülikoolile tasuta loa (lihtlitsentsi) enda loodud teose

Eesti framenet'i viimine MySQL'i

mille juhendaja on Neeme Kahusk

reprodutseerimiseks säilitamise ja üldsusele kättesaadavaks tegemise eesmärgil, sealhulgas digitaalarhiivi DSpace-is lisamise eesmärgil kuni autoriõiguse kehtivuse tähtaja lõppemiseni;

üldsusele kättesaadavaks tegemiseks Tartu Ülikooli veebikeskkonna kaudu, sealhulgas digitaalarhiivi DSpace'i kaudu kuni autoriõiguse kehtivuse tähtaja lõppemiseni.

olen teadlik, et punktis 1 nimetatud õigused jäävad alles ka autorile.

kinnitan, et lihtlitsentsi andmisega ei rikuta teiste isikute intellektuaalomandi ega isikuandmete kaitse seadusest tulenevaid õigusi.

Tartus 14.05.2014